```
from google.colab import drive
drive.mount('/content/drive')
```

    Mounted at /content/drive

```
cd /content/drive/MyDrive/CSE475/project1
```

    /content/drive/MyDrive/CSE475/project1

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import accuracy_score
```

```
dataset = pd.read_csv('covid_dataset.csv')
dataset
```

|	Day	Lab Test	Confirmed case	Death Case
0	2020-04-04	434	9	2
1	2020-04-05	367	18	1
2	2020-04-06	468	35	3
3	2020-04-07	679	41	5
4	2020-04-08	981	54	3
...	...	...	...	...
621	2021-12-16	25203	257	3
622	2021-12-17	16310	191	2
623	2021-12-18	13991	122	4
624	2021-12-19	19332	211	1
625	2021-12-20	19955	260	2

626 rows × 4 columns

```
x= dataset [['Lab Test']]
y= dataset[['Confirmed case']]
x_train , x_test ,y_train,y_test = train_test_split (x,y,test_size = 0.3 , random_state = 42
```

```
model = LinearRegression()
model.fit(x_train,y_train)
```

```
    LinearRegression()
```

```
plt.xlabel('Lab Test')
plt.ylabel('Confirmed case')
plt.scatter(x_train,y_train, color='g')
plt.plot(x_train, model.predict(x_train),color='r')
plt.show()
```
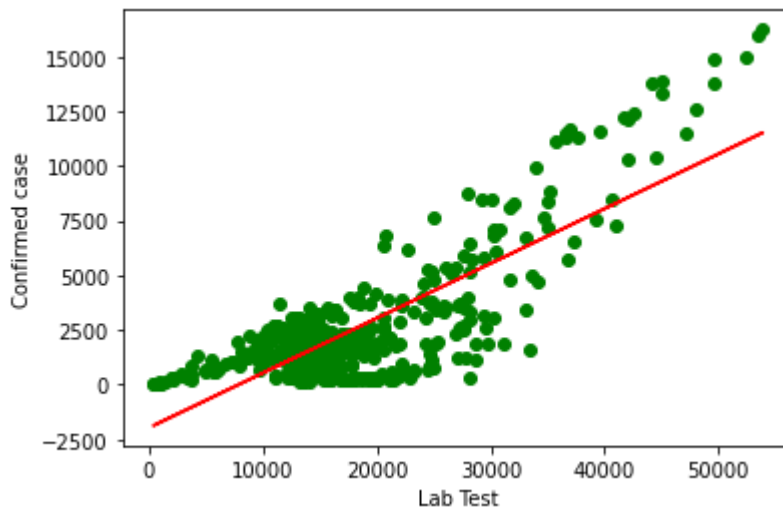


```
model.score(x_test,y_test)
```

```
poly = PolynomialFeatures(degree=2, include_bias=False)
poly
```

```
    PolynomialFeatures(include_bias=False)
```

```
x_poly = poly.fit_transform(x.values.reshape(-1, 1))
```

```
model2 = LinearRegression()
```

```
model2.fit(x_poly, y)
```

```
    LinearRegression()
```

```
y_predicted = model2.predict(x_poly)
y_predicted
```

```
    array([[ 1272.09531879],
           [ 1278.3915804 ],
```

```
       [ 1268.92526764],
       [ 1249.62965395],
       [ 1223.14312693],
       [ 1229.68323331],
       [ 1206.08758377],
       [ 1225.45692846],
       [ 1193.3896655 ],
       [ 1175.31643068],
       [ 1150.3739955 ],
       [ 1162.45432286],
       [ 1142.25975732],
       [ 1130.4441441 ],
       [ 1135.64282493],
       [ 1101.75822588],
       [ 1093.01349584],
       [ 1081.73730247],
       [ 1074.96475647],
       [ 1058.23315116],
       [ 1045.2786791 ],
       [ 1062.2247899 ],
       [ 1055.26240514],
       [ 1039.59747765],
       [ 1018.60331765],
       [  998.29259835],
       [  998.37454177],
       [  984.45241312],
       [  980.23448475],
       [  988.54349874],
       [  975.21548674],
       [  982.04393139],
       [  975.37830157],
       [  979.65608689],
       [  978.64762994],
       [  986.53124657],
       [  983.2833469 ],
       [  973.78350711],
       [  972.81165305],
       [  981.02234709],
       [  975.0259809 ],
       [  994.99676367],
       [  972.80377387],
       [  986.01652174],
       [ 1036.32869463],
       [  991.73855556],
       [ 1055.65871546],
       [ 1058.38636856],
       [ 1033.72826997],
       [ 1089.37190421],
       [ 1004.07531058],
       [ 1022.64136081],
       [  987.71793729],
       [  982.90276917],
       [ 1017.40658858],
       [ 1118.211377  ],
       [ 1045.18972645],
       [ 1158.09389523],
```

```
plt.figure(figsize=(10, 6))
plt.xlabel('Lab Test')
plt.ylabel('Confirmed case')
plt.title("daily new covid patient case ", size=16)
plt.scatter(x, y,c="g")
plt.plot(x, y_predicted, c="red")
plt.show()
```
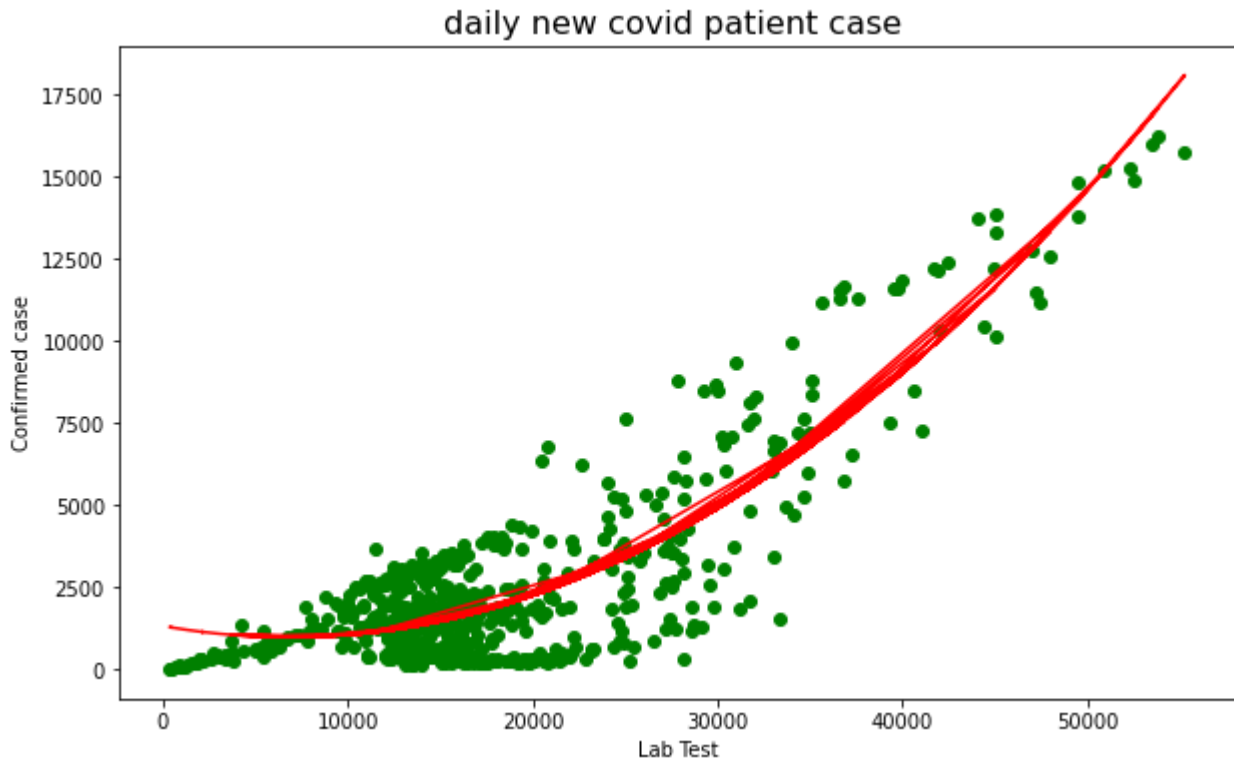


```
model2.score(x_poly,y)
```

```
0.7667815388703842
```

Conclusion

Here, using linear regression, the accuracy is 66%. On the other hand, when we use polynomial regression, we achieve 76% accuracy.